CS230

Transfer Learning Based Multi-fidelity Surrogate Model for Lithium-ion Battery Pack

Wenting Ma
Department of Energy Science & Engineering
Stanford University
wma21@stanford.edu

Su Jiang*
Department of Energy Science & Engineering
Stanford University
sujiang@stanford.edu

Abstract

Physics-based models are important tools in lithium-ion battery research and development. Yet their computational cost is impractically high for simulating commercial-size battery packs consisting of hundreds of cells. Here we propose a CNN-LSTM battery surrogate model used for uncertainty quantification based on transfer learning and multi-fidelity simulation data from a physics-based simulation model. First, a low-fidelity surrogate model is trained on a large single-cell dataset. Then, we freeze the lower LSTM layers and retrain the upper LSTM layer using medium-size small pack datasets of various pack configurations. Finally, we freeze all LSTM layers and retrain the input and output layers using a small 3s37p (3 series 37 packs) pack dataset. The model achieves a percentage RMS error of 1.08%, compared to 2.17% for a directly-trained model without transfer learning.

1 Introduction

In lithium-ion battery design, operation, and prognosis, physics-based models show irreplaceable advantages in prediction accuracy and adaptability. When utilizing physics-based battery pack models, it is common practice to assume electrochemical parameters for each battery cell within the pack to be identical at the nominal value. However, in reality, cell parameters are subject to manufacturing variations and non-uniform degradation rates due to uneven branch currents and temperature gradients. This parameter heterogeneity within the pack leads to uncertainty in model-predicted battery pack performance, such as pack voltage, capacity, and average temperature.

In this project, we quantify the uncertainty in physics-based model prediction of battery pack terminal voltage. Due to high computational cost associated with Monte Carlo sampling on physics-based simulation, we constructed a CNN-LSTM surrogate model for Li-ion battery packs based on transfer learning and multi-fidelity simulation data. The input of our algorithm are design parameters of all cells within a battery pack. Specifically, there are 13 uncertain model parameters and 2 pack configuration parameters for each cell, so for a pack containing 9 cells, the input of one data sample has the shape of $(15 \times 9,)$. We then use a neural network model with transfer learning techniques to output a time series of predicted pack voltages. We also output a time series of binary flags indicating whether the battery pack has finished its cycle profile.

*Su Jiang currently holds a post-doc position at Stanford University and she is not enrolled in CS230 class.

CS230: Deep Learning, Fall 2022, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

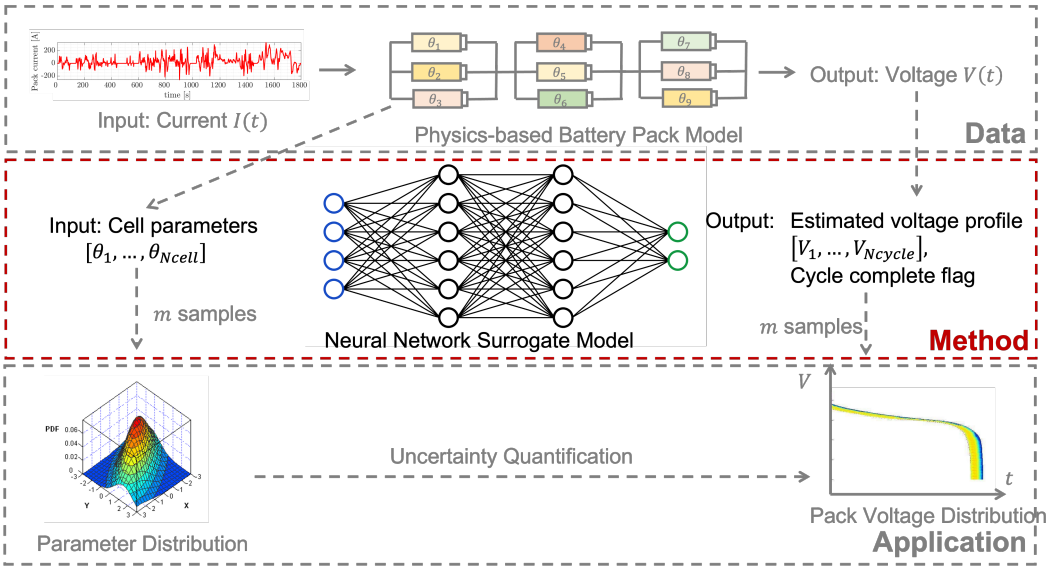


Figure 1: Lithium-ion Battery Electrochemical Model Uncertainty Quantification Workflow Overview

2 Related work

There are many studies focusing on uncertainty quantification (UQ) of single cell battery model, using Monte Carlo Simulation [1], polynomial chaos expansion (PCE) [2], polynomial response surfaces, kriging, and radial-basis neural networks (RBNN) [3]. The common motivation of adopting alternative UQ methods is to reduce the high computational cost of battery model simulation. Still, UQ using PCE in [2] with 17 uncertain parameters requires 1000 simulations to quantify cell capacity uncertainty under a constant-current discharge process. The UQ work in [3] using kriging and RBNN with 3 uncertain parameters select 315 simulation points in the parameter space to build the surrogate models. The computational cost of simply applying these methods on a physics-based battery pack model consisting of hundreds of cells would be impractically expensive.

Application of neural network surrogate models in lithium-ion batteries ranges from UQ[4] to degradation prognosis[5] and state-of-charge estimation[6]. [4] focuses on bi-fidelity transfer learning for UQ, where 17 uncertain battery parameters are fed into a feed-forward neural network (FNN) and a residual neural network (ResNet) respectively to fit the model-predicted end-of-cycle liquid phase concentration. The work refers to coarse and fine discretization grids in simulation as high and low fidelity data. [6] constructed an LSTM surrogate model for online battery state-of-charge estimation. The surrogate model achieves 135 times faster computational speed at the cost of 2% mean absolute error compared with 0.72% for original physics-based model. The work presents promising efficacy of LSTM as a battery surrogate model with sequential outputs.

To the best of our knowledge, no study has been conducted to utilize multi-fidelity data for surrogate modeling of physics-based battery pack models.

3 Dataset and Features

We use synthetic data generated from battery pack enhanced single-particle model [7] under a constant-current cycling profile to train the neural network surrogate model. Table. 1 explains the number of available data samples under each type of pack configuration. Note that 3s3p and 3s37p are two representative cases where accuracy of output distribution is additionally examined, apart from accuracy of point estimates. Therefore, in these two cases 1000 data samples are set aside as "ground truth output distribution" for testing.

The input features of each data sample are design parameters of all cells within a battery pack. For each cell, 13 uncertain parameters are independently and randomly generated subject to given distributions. This parameter vector is then concatenated with 2 pack configuration parameters indicating the location of the cell within the pack (explained in Sec 8.1).

Pack Configuration	# of samples	Train-val-test split
1s1p	1000	(800, 100, 100)
6s1p	200	(160, 40, 0)
1s6p	200	(160, 40, 0)
3s3p	200	(160, 40, 1000)
3s6p	200	(160, 40, 0)
3s37p	100	(80, 20, 1000)

Table 1: Dataset overview across different battery pack configurations. "XsYp" refers to a battery pack with Y parallel-connected cells for each group and X groups of cells connected in series.

The output pack voltage profiles are sampled at 1/60 Hz with uniform length of 83 minutes. We performed zero-padding at the end of profiles shorter than 83 minutes, and introduced a binary flag concatenated with pack voltage at each time step as output to indicate whether the battery has reached the end of its cycle profile.

Input features are normalized by subtracting the mean of each feature and dividing by its standard deviation. Output voltage profiles are normalized using their mean and variance at each time step across all training samples. Examples of input and output data are visualized in Sec 8.2 in the Appendix.

4 Methods

4.1 Baseline

To validate the pack-level performance of proposed transfer learning models, four baselines are selected for comparison:

- Average (Avg): Average voltage profile across all training examples from the same pack.
- Scaled Single-cell (Scale): Ground truth voltage of 1s1p pack, i.e. single cell, scaled according to pack configuration. Specifically, to examine the performance of a N_ssN_pp pack, the baseline scales the ground truth single-cell voltage V_{1s1p} as:

$$V^{SS} = N_s * V_{1s1p} \tag{1}$$

  Note that input features of single-cell samples and pack samples are of different shapes and values, so the baseline is only used for distribution analysis.
- Pack Without Interconnection (Packw/o): For a pack sample, simulate each cell in the physics-based model with corresponding design parameter without any thermal or electrical interconnection, then use simple rules to estimate pack voltage. The baseline aims at verifying whether the proposed model captures cell-to-cell interaction dynamics.
- Direct Training (DT): Neural network model of same architecture trained only on high-fidelity data samples.

4.2 Bi-fidelity Transfer Learning

We propose a transfer learning scheme using bi-fidelity data based on a Long Short-Term Memory (LSTM) model. Model architecture and detailed transfer learning steps are illustrated in Sec 8.3. The goal is to estimate 3s3p pack voltage profile. An LSTM layer with 8 filters is applied as an encoding layer to map the input from different fidelity data to the same dimension and extract latent features. The outputs of the encoding layers are fed to the decoder composed of two LSTM layers, each consisting of 32 filters. We choose this architecture because LSTM enables neural networks to learn both short-term and long-term dependencies within the data.

The objective is to minimize the mean-squared-error (MSE) between true and predicted voltages before the end-of-cycle for each sample, while minimizing the binary cross-entropy loss for end-of-cycle flags. Denote time length of output profile as T, the ground truth output as $\mathbf{Y} = [\mathbf{V}, \mathbf{b}]$, and the model-estimated output as $\hat{\mathbf{Y}} = [\hat{\mathbf{V}}, \hat{\mathbf{b}}]$, where $\mathbf{V}$ is a $T \times 1$ vector of pack voltage, and $\mathbf{b}$ is a $T \times 1$ vector of end-of-cycle flags. The loss function is

$$Loss = \frac{\sum_t (1 - b_t)(V_t - \hat{V}_t)^2}{\sum_t (1 - b_t)} - \frac{\alpha}{T} \sum_t (b_t * \log(\hat{b_t}) + (1 - b_t) * \log(1 - \hat{} \, b_t)) \tag{2}$$

where α is a weight hyperparameter indicating penalty imposed on end-of-cycle flag prediction.

4.3 Multi-fidelity Transfer Learning

A significant downside of the LSTM model architecture proposed in Sec 4.2 is the exploding number of trainable model parameters in Step 2 and 3 with the increase of pack size. Table 3 in Sec 8.4 shows the number of trainable model parameters in transfer learning and computational cost to obtain 200 samples for different pack sizes. Under the same computational budget, the previous LSTM model architecture faces overfitting issue for commercial-size battery packs such as 6s74p (on Tesla Model S) due to lack of training samples.

To address the overfitting issue on small-dataset, we propose a second transfer learning scheme using multi-fidelity data based on a CNN-LSTM model, as shown in Figure 2. The low-fidelity (LF) dataset includes $N_l = 800$ single cell samples. Mid-fidelity (MF) data are small pack samples for 1s6p, 6s1p, 3s3p, and 3s6p. There are $N_m = 160$ training samples for each pack configuration. High-fidelity (HF) data are $N_h = 160$ 3s37p pack samples. The goal is to estimate 3s37p pack voltage profile. A CNN encoding layer followed by a max-pooling layer is applied in replace of LSTM encoding to map the input to the same dimension and extract salient latent features over large-size input parameters. Table 4 in Sec 8.4 explains CNN kernel size selection under each pack configuration. The outputs of the encoding layers are fed to the decoder of the same architecture as in Sec 4.2, then fed to another LSTM layer with 16 filters.

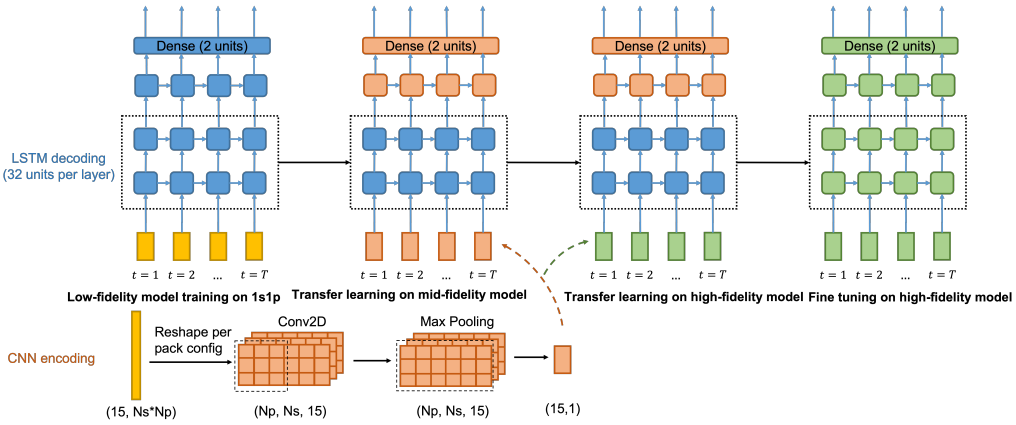


Figure 2: Multi-fidelity transfer learning scheme.

As shown in Figure 2, we first train a low-fidelity model using single cell samples so that lower LSTM layers learn cell-level physics. LF data share the same cell-level physics as MF and HF data, so estimated cell voltages from LF model after scaling are close to actual pack voltages. Still, we need high-fidelity pack-level data to capture cell-to-cell thermal and electrical interaction dynamics. We then freeze the weights in LSTM decoder, and retrain the CNN encoding layer and upper LSTM layer using various small pack samples. Finally, we use large pack samples simply to train the CNN encoding layer, and fine-tune the entire HF model using a small amount of HF data.

5 Results and Discussion

Results of bi-fidelity transfer learning on a 3s3p pack are presented in Sec 8.5. Here we illustrate the performance of multi-fidelity transfer learning model on a 3s37p pack. We select penalty coefficient $\alpha = 0.05$ in loss function given in Eq. 2. The hyperparameter search process for α is detailed in Tab 6 in Sec 8.4. We select initial learning rate at 0.01 for low-fidelity and mid-fidelity model training, and use keras `ReduceLROnPlateau` to reduce learning rate when training loss stops improving. The maximum number of epochs is 200. The batch size is 20 for single cell and 4 for packs. For fine-tuning of high-fidelity model, the learning rate starts at 3e-4 and decreases similarly, and the maximum number of epochs is 100. This choice of hyperparameters ensures reasonable model training time within 30 minutes (in cpu) without sacrificing model performance.

We use percentage RMS as evaluation metric for point-estimates:

$$percentage\ RMS = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(V_t - \hat{V}_t)^2} \times \frac{N}{\sum_{t=1}^{T} V_t} \times 100\% \tag{3}$$

where V_t and $\hat{V}_t$ are ground truth and predicted voltage respectively.

The percentage RMS of our method compared with baselines is shown in Table 2. Fig. 3 illustrates 5 validation samples of predicted voltage profiles. The proposed model achieves lower point-estimate error compared with all baselines. Specifically, Baseline DT tends to overestimate end-of-cycle voltage in all cases due to overfitting. Baseline Packw/o tends to underestimate voltage in the later half of cycle (charging phase). In contrast, the proposed model manages to capture the general trend of cell-to-cell interaction dynamics.

Dataset	Proposed	Avg	Packw/o	DT
Validation	1.17	-	1.26	2.30
Test	1.08	2.13	-	2.17

Table 2: Percentage RMS (%) of proposed method compared with baselines.

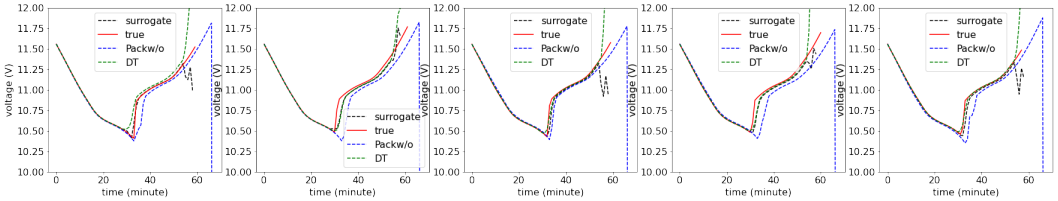


Figure 3: Examples of predicted voltage profiles ("surrogate") compared with baselines.

Fig. 4 shows model performance in terms of predicted distribution across 1000 test samples. Table 5 in Sec 8.4 shows the Wasserstein Distance of predicted end-of-cycle voltage and time distribution comapred with baselines. Overall, the proposed method is closer to actual distribution in terms of entire voltage profile, end-of-cycle voltage, and end-of-cycle time.

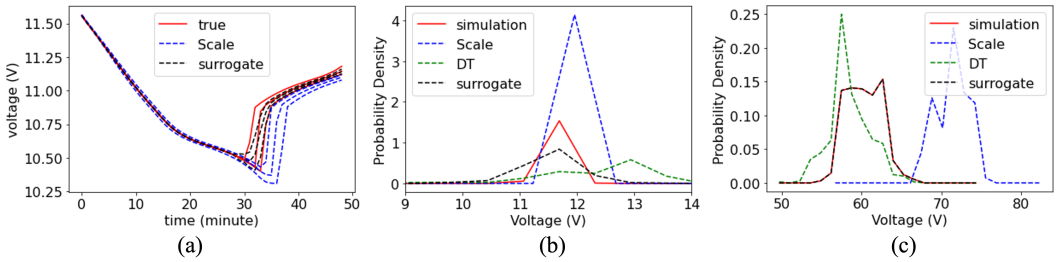


Figure 4: Multi-fidelity transfer learning distribution performance. (a) 10-50-90 percentiles of voltage profiles. (b) Distribution of end-of-cycle voltage. (c) Distribution of time.

6 Conclusion and Future Work

In this project, we proposed a multi-fidelity transfer learning scheme for surrogate modeling of commercial-size battery packs to address the over-fitting issue observed in bi-fidelity transfer learning. The proposed CNN-LSTM model significantly prevents overfitting and captures cell-to-cell interaction dynamics within the battery pack. Future work includes leveraging the proposed multi-fidelity learning scheme to train battery surrogate models under real-world cycles and under multiple cycles. This surrogate model will serve as a fast prediction tool for efficient uncertainty quantification in my future research.

7 Contributions

The idea of performing UQ on battery pack is proposed by Professor Simona Onori. Su Jiang advised on the design of LSTM model architecture and provided the initial version of bi-fidelity transfer learning codes. Wenting Ma completed the rest of the work in the project independently, including data generation and preprocessing, modification on bi-fidelity transfer learning architecture, design and implementation of multi-fidelity transfer learning, hyperparameter tuning and architecture exploration, creating visualizations, and writing all reports.

8 Appendix

8.1 Computation of location parameters in input features

From a physical-model perspective, the location of each battery cell in the pack is important in battery performance estimation because it determines cell operation temperature under certain coolant flow, and all cells within one parallel-connected group have electrical constraints on each other. For the cell in i-th cell group and at j-th parallel branch, its location parameters, denoted as $\mathbf{x}_{loc}$, in a $N_s s N_p p$ pack are calculated as:

$$\mathbf{x}_{loc} = [\frac{i}{N_s}, \frac{j}{N_p}] \tag{4}$$

8.2 Dataset Visulization

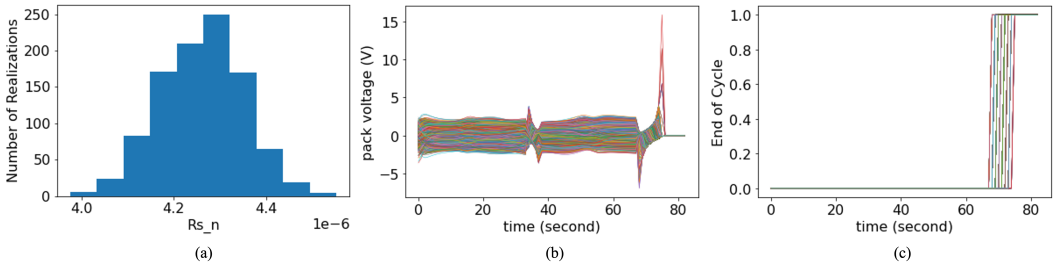


Figure 5: (a) Distribution of input design parameter $R_{s,n}$, i.e. anode particle radius, across 1000 single-cell data samples. (b) Examples of normalized output voltage profiles. (c) Examples of output end-of-cycle flags.

8.3 Bi-fidelity Transfer Learning Scheme

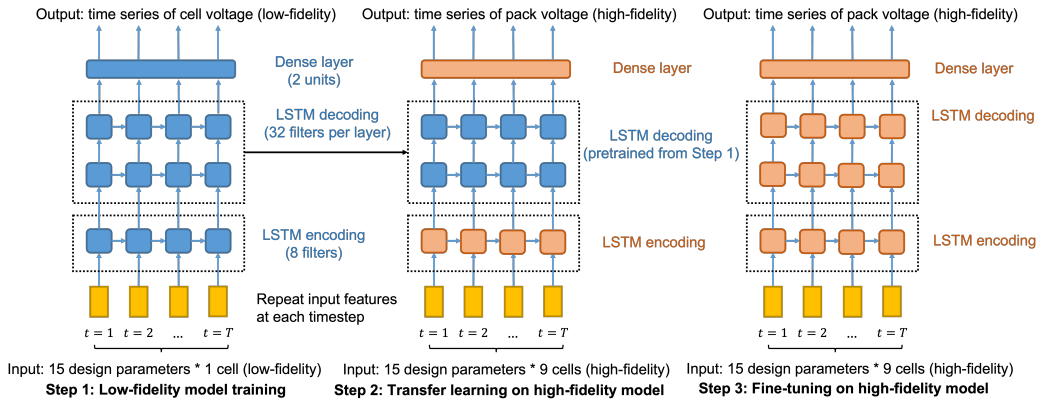


Figure 6: Bi-fidelity transfer learning scheme using single-cell data (LF) and 3s3p pack data (HF).

Pack Config	Required Time (s) [2]	# of Trainable Parameters outside
1s1p	556.14	-
3s3p	1011	1698
3s37p	10485	13938
6s74p	45024	53898

Table 3: Computational cost and number of trainable parameters in LSTM model in Sec 4.2 for various pack configurations.

8.4 Tables

Pack Config	CNN Kernel Size	MaxPooling Pool Size
1s6p	(3,1)	(6,1)
6s1p	(1,3)	(1,6)
3s3p	(3,3)	(3,6)
3s6p	(3,3)	(6,3)
3s37p	(3,3)	(37,3)

Table 4: CNN Kernel Size Selection

Predicted Variable	Proposed	Scale	DT
End-of-cycle voltage	0.0435	0.0834	0.05597
End-of-cycle time	0	0.0132	0.0252

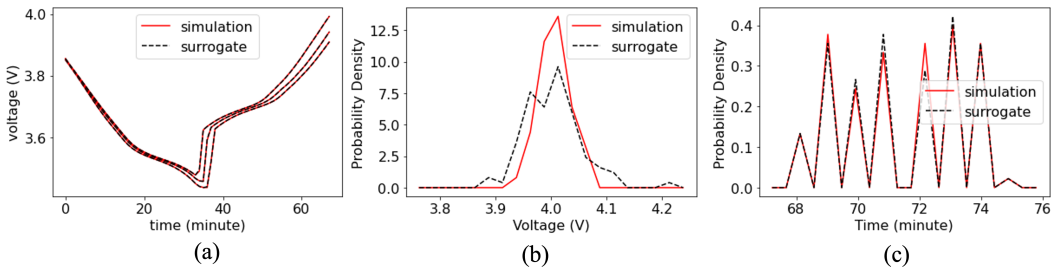
Table 5: Wasserstein Distance of predicted end-of-cycle voltage distribution and time distribution for a 3s37p battery pack.

α	0.01	0.05	0.1	0.5	1
Test Percentage RMS Error (%)	0.30	**0.11**	0.19	0.20	0.37

Table 6: Hyperparameter search of penalty coefficient α in loass function Eq. 2 on the single-cell model.

8.5 Single Cell and Bi-fidelity Model Performance

Percentage RMS for single-cell surrogate model across 100 test samples is 0.19%. Percentage RMS for 3s3p surrogate model across 1000 test samples is 2.05% compared with 6.07% for Baseline Avg.



(a) (b) (c)

Figure 7: Single cell distribution performance. (a) 10-50-90 percentiles of voltage profiles. (b) Distribution of end-of-cycle voltage. (c) Distribution of time.

[2]The required time is obtained on a local CPU.

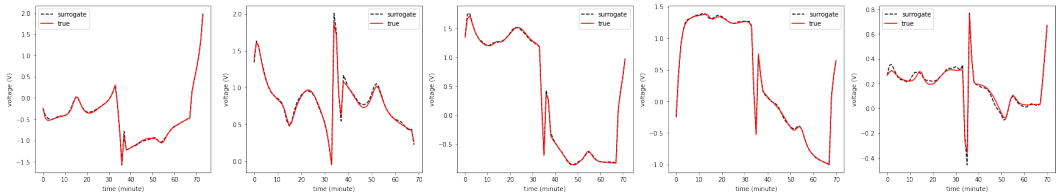
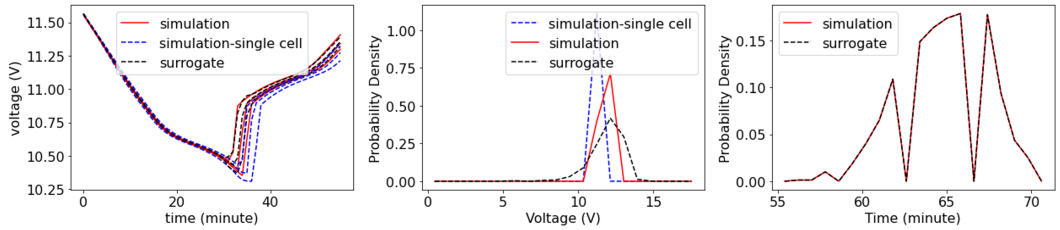


Figure 8: Examples of predicted voltage profiles for single cell compared with baselines.



(a) (b) (c)

Figure 9: Bi-fidelity transfer learning distribution performance. (a) 10-50-90 percentiles of voltage profiles. (b) Distribution of end-of-cycle voltage. (c) Distribution of time.

References

[1] W. Tong, W. Q. Koh, E. Birgersson, A. S. Mujumdar, and C. Yap, "Correlating uncertainties of a lithium-ion battery–a monte carlo simulation," *International Journal of Energy Research*, vol. 39, no. 6, pp. 778–788, 2015.

[2] M. Hadigol, K. Maute, and A. Doostan, "On uncertainty quantification of lithium-ion batteries: Application to an lic6/licoo2 cell," *Journal of Power Sources*, vol. 300, pp. 507–524, 2015.

[3] W. Du, A. Gupta, X. Zhang, A. M. Sastry, and W. Shyy, "Effect of cycling rate, particle size and transport properties on lithium-ion cathode performance," *International Journal of Heat and Mass Transfer*, vol. 53, no. 17-18, pp. 3552–3561, 2010.

[4] S. De, J. Britton, M. Reynolds, R. Skinner, K. Jansen, and A. Doostan, "On transfer learning of neural networks using bi-fidelity data for uncertainty propagation," *International Journal for Uncertainty Quantification*, vol. 10, no. 6, 2020.

[5] P. Li, Z. Zhang, R. Grosu, Z. Deng, J. Hou, Y. Rong, and R. Wu, "An end-to-end neural network framework for state-of-health estimation and remaining useful life prediction of electric vehicle lithium batteries," *Renewable and Sustainable Energy Reviews*, vol. 156, p. 111843, 2022.

[6] S. Oh, K. Jang, J. Kim, and I. Moon, "Online state of charge estimation of lithium-ion battery using surrogate model based on electrochemical model," in *Computer Aided Chemical Engineering*, vol. 51, pp. 1447–1452, Elsevier, 2022.

[7] T. Weaver, A. Allam, and S. Onori, "A novel lithium-ion battery pack modeling framework-series-connected case study," in *2020 American control conference (ACC)*, pp. 365–372, IEEE, 2020.